DOCUMENT RESUME

ED 064 402 TM 001 650

AUTHOR Pruzek, Robert M.; And Others
TITLE On the Analysis of Partitioned Data.
NOTE 17p.

EDRS PRICE MF-$0.65 HC-$3.29
DESCRIPTORS *Cluster Grouping; *Data Analysis; Data Collection; Hypothesis Testing; Organization; *Research Methodology; *Statistics

ABSTRACT

A description is given of a general method for studying partitions. The main focus is with the analysis of relationships among several different partitions of the same items for the explorations as well as confirmation of structural relationships. A partition is defined as a set of mutually exclusive clusters of items; however, this paper deals with overlapping or conjoint clusters also. The terms "clustering" and "partitioning" are found in the context of both data collection and analysis. The major impetus for the present method was the problem of studying how an individual (manifest) partition might be examined in relation to a single derived (latent) partition in the context of latent partition analysis (LPA). In the process of studying characteristics of the distribution of the principal statistic, there was also developed an approach to the organization of a set of partitioned data with respect to any specified target. Data were analyzed first with respect to an a priori target which had been generated by the investigators on the basis of characteristics of the Morse Code symbols, with no regard for the letters themselves. It was found that: (1) The organization and display features of the proposed strategy may be of greatest significance, especially when stimuli or items are complex and when sorters are heterogeneous in terms of how they formed their categories; and (2) There are several other possible indices of association for comparing matrices. (CK)

7.9

ED 064402

ON THE ANALYSIS OF PARTITIONED DATA

Robert M. Pruzek
State University of New York at Albany

Charles A. Stegman
University of Pittsburgh

Robert A. Pfeiffer
State University of New York at Albany

Whenever anyone studies a variety of items, objects or variables it is almost inevitable that at some stage he will cluster them into categories. Such clustering seems to be a fundamental part of any search for knowledge since it occurs whenever we name objects or distinguish among different classes of things. It is unnecessary to document here that clustering procedures have been used in practically every area of science and that a wide variety of methods have been employed. Yet, there have been few systematic reports on the characteristics of methods for clustering or partitioning data. The purpose of this paper is to describe a general method for studying partitions which appears to us to have special potential for research in the behavioral and social sciences. Our major concern is with the analysis of relationships among several different partitions of the same items for the exploration as well as confirmation of structural relationships. In this paper, a partition is defined as a set of mutually exclusive or disjoint clusters of items but our approach may be rather

TM 001 650

simply extended to deal with overlapping or conjoint clusters. Clusters within a single partition are taken to have no quantitative ordering; such clusters are considered as merely qualitatively distinct.

The terms "clustering" and "partitioning" are to be found in the context of both data collection and analysis. It may be helpful to provide some methodological perspective on various uses for these terms, and on the variety of methods which have been developed. Perhaps the most common use of the term "clustering" has been in the context of data analysis where N objects or variables have been used to generate all of the N(N-1)/2 pairwise measures of association from which some type of dimensional analysis is initiat d. The methods of Hartigan (1967) for tree structure analysis, Johnson (1967) for hierarchical clustering and the Guttman (1968) Kruskal (1964 a,b) and Shepard (1962) approaches to multidimensional scaling all require initially such pairwise measures of relatedness (e.g., proximity, similarity, distance, correlation). Closely related in certain ways are the methods of parametric mapping (Shepard and Carroll, 1966) and multidimensional unfolding (Coombs, 1950; Barnett and Hays, 1960). Kruskal (1969), incidentally, has reviewed such methods and provided some new insights as to their similarities and differences. Harris and Kaiser (1964) and Tryon (1958) have both used the term "clustering" in the context of factor analysis, although they have used the word in somewhat different ways. Hartigan has also begun work on simultaneous clustering of rows and columns of rectangular score matrices; while yet unpublished, this may be an important new direction for clustering methods since it involves a direct approach to organizing and understanding relationships among observations. Some additional sources have also been included in our References to facilitate further examination of the literature.

The terms "partitioning" and "clustering" have often been used interchangeably in the context of data collection and analysis. Wiley's (1967) latent partition analysis (LPA) defines "manifest partitions" at the level of data collection as well as "latent partitions" at the level of analysis. LPA has served to stimulate a number of studies involving further elaboration and refinement of partitioning methodology (see INDEX - SIG Sort, Archives and Evans (1970)). In fact, the work reported here was motivated in part by what were seen as problems or limitations with the LPA approach.

The major impetus for the present method was the problem of studying how an individual (manifest) partition might be examined in relation to a single derived (latent) partition in the context of LPA. But we began to find ourselves interested in a more basic problem in the analysis of partitions. Instead of examining relationships among individual partitions of a set of items and some type of average partition such as the type which is produced in LPA applications, we decided that a means should be sought of measuring the goodness (or badness) of fit of any single partition of N items to what we shall call a "target" partition. Such a measure was devised and it has been found to be useful for both the exploration and confirmation of structural relationships among a set of N items. In the process of studying characteristics of the distribution of our principal statistic, we have also developed what we see to be a useful approach to the organization of a set of partitioned data with respect to any specified target. The remainder of this paper is to describe details of the proposed method and to apply it in the analysis of data.

Suppose, initially, that we have N items which have been partitioned into n_s categories by each of S sorters ($s=1, \ldots, S$). For each sorter we may construct a binary matrix of order N x N the entries of which index joint occurances (1), and non-occurance (0), of the various item pairs for this sorter. Let us call such a matrix A_s.

Next, consider a binary matrix labeled A_t, which is analoguous to A_s, except that it has been constructed as a model or "target" partition for the set of N items. A_t is also a joint-occurance matrix but it is taken as fixed; it might typically correspond to an experimenter's hypothesis about some "cue" system which sorters have used in partitioning the items. Ultimately, we shall also consider the possibility that any number, T, of targets may have been specified a priori so that the A_t could range over $t=1, \ldots, T$.

Also let $k_1, \ldots, k_{n_s}$ be the frequencies of items in the sth sorter's partition and $m_1, \ldots, m_{n_t}$ be the corresponding frequencies for the tth target. Later we shall have occasion to use functions defined from these numbers as, $h(A_s) = \frac{\Sigma k_i(k_i-1)}{N(N-1)}$ and $h(A_t) = \frac{\Sigma m_i(m_i-1)}{N(N-1)}$ where these two quantities may be called the partition heights.

When one's object is to study a set of partitions in relation to one or more targets, it becomes useful to assess the degree of association between A_s and A_t using a summary statistic. One such statistic which has been found efficacious is the quantity

$$q_{st} = \text{tr } (A_s - A_t)^2/N(N-1) \qquad (1)$$

where the numerator on the right-hand side represents the total number of discrepancies between A_s and A_t. Since all diagonal elements of both matrices are necessarily unity, it follows that N(N-1) is generally the maximum value of the tr ( ); thus q_{st} might be regarded as a normalized trace which ranges between 0 and 1. q_{st} is directly interpretable as a badness of fit statistic for the pair (A_s, A_t); the statistic is symmetric in that permutations of (A_s, A_t) to (A_t, A_s) do not change its value; and it is easily computed as the proportion of disagreements between A_s and A_t. Also, it can be shown that

$$q_{st} = h(A_s) + h(A_t) - 2h(A_s \cap A_t) \qquad (2)$$

where $(A_s \cap A_t)$ corresponds to the element-wise product, or intersection, of the pair (A_s, A_t). A goodness of fit statistic might be written as $p_{st} = 1-q_{st}$.

Substantial investigation has already been made of the distributional properties of q_{st}. It has been possible to write a computer program to generate the exact distribution of q_{st} for any final target where $h(A_s)$ is also fixed for A_s. Combinational equations are used to generate all possible outcomes for A_s for any A_t; the program may then be used to compute each of the corresponding values of q_{st}. The expectation of q_{st} has been found to be particularly simple:

$$E(q_{st}) = h(A_s) + h(A_t) - 2h(A_s)\, h(A_t) \qquad (3)$$

Further, more detailed results for the exact distribution are not included here, however. This is because the computer time required to find the exact distribution of q_{st} actually exceeds the time required for moderately large Monte Carlo distribution of q_{st}. Moreover, q_{st} has been found to be virtually Gaussian as N grows to 15 or 20 items or more for most "realistic" $h(A_s)$, $h(A_t)$ combinations. Tables 4a and 4b include two such Monte Carlo distributions of q_{st}. As N grows above N = 20 for fixed heights of A_s and A_t, the standard error of q_{st} will necessarily decrease for fixed heights, $h(A_s)$, $h(A_t)$.

In order to apply the proposed method for the analysis of data, an experiment was designed in which 50 graduate students were asked to partition a set of 26 items, each of which was a pairing of a letter from the alphabet and its own Morse Code equivalent. Free sort instructions were used, with the only stipulation being that each student should form categories of items which in some sense might be thought of as mutually homogeneous. Sorters were thus allowed to choose their own system for partitioning; the intention was to help insure that several different bases would be chosen for the partitions. All students were asked to use no less than five, or no more than nine, categories.

The data were analyzed first with respect to an a priori target which had been generated by the investigators on the basis of characteristics of the Morse Code symbols, with no regard for the letters themselves. Thus, the aim was to test the hypothesis for each individual that his partition is random with respect to the target; this is the null hypothesis. Rejection of this hypothesis can be taken as constituting evidence, at some specified probability level, that the partition in question may not reasonably be regarded as randomly different from the target. To have a sufficiently small value of q_{st} (or large value of p_{st}) is taken as evidence that the target in question can reasonably be regarded as having been the model in some sense for the individual's manifest partition. As with all procedures for inductive inference the interpretation should be qualified in that both Type I and Type II errors are possible.

Table 1, is a display of the original data, organized by columns and rows with respect to target 1. Columns, representing items, were simply grouped according to the eight a priori target clusters. Rows, corresponding to sorters, were ranked, first by the number of categories which had been formed and, next, by using the statistic q_{st} = normalized tr $(A_s - A_t)^2$. Also, the height coefficients were calculated from the respective sets of category frequencies for the sorters. For each sorter the nominal category numbers for the original partitions were then printed.

These results show that five of the sorters (Nos. 30, 42, 46, 48, and 18) apparently used the same cues for their partitions, as were used in generating this target; as expected, the associated q_{st} (or trace) statistics are zero. Moreover, a number of other sorters had minimal confusions with respect to this target. For individuals with eight or nine category partitions, the theoretically expected value of the q_{st} statistic is approximately .18. Since the sampling distribution of q_{st} tends to have only mild negative skewness (for eight category partitions with "small" heights) and the estimated standard error (from Monte Carlo data) of q_{st} = .04, it might be argued that persons with q_{st}'s $\leq .10$ have not randomly sorted with respect to the target but, rather, have in some sense used the model of the target partition. Inspections of the individual sorters partitions are readily made using tables such as this one.

Further analyses were conducted using another target which was constructed from the results of an LPA analysis. Table 2 includes a contingency table with which the a priori and the LPA (target) partitions may be compared. Table 3 is an analogue of Table 1 for the LPA-derived target, which has been labeled Target 2.

As might be expected, the q_{st} values tend generally to be smaller for Target 2 than Target 1 ($\bar{q}_{.1}$ = .128, $\bar{q}_{.2}$ = .123), but it remains interesting that none of q_{s2} values are identically zero. Despite its overall similarity to this set of sorter partitions, the derived LPA partition does not seem precisely to have been any single person's model in the sense that Target 1 apparently was. Again, further examinations of the rows of Tables 1 and 3 might be useful for further exploration of the data; contingency tables such as that of Table 2 might also facilitate more detailed study.

Were we to have specified other target partitions with respect to manifest partitions of these 26 alphabetic-Morse-Code combinations, the data could of course just as easily have been organized and displayed for each target. We suggest that for complex items, generally, where several different schemes might have been employed to generate manifest partitions, that detailed analysis using multiple targets can provide an efficient and thorough analysis of one's basic data. Comparisons of statistics of the form of q_{st}, or vectors of the form of $\vec{q}_{s.}' = q_{s1}\ q_{s2} \cdots q_{st}$, with other measures at the level of the sorters (or groups of sorters) can be developed as direct analogues of standard methods for analysing quantitative data.

Finally, it might be helpful to make some further, more general points, about the suggested strategy for studying partitions in relation to other methods which are available. Wiley's (1967) latent partition analysis and Johnson's (1967) hierarchical clustering have both proven especially useful for explorations of partitions. Although the present method might also be used for exploration as well as explanation, we see the new approach as complementing, not competing with, the earlier ones. As we have shown above, the earlier methods may be used to generate targets and hence, to organize the set of partitions. The new strategy, based on more direct analysis of the partitions themselves, however, does seem to have more potential for refining one's understanding of the bases which have been used for sorting; there seems to be a substantial advantage for interpretation, and sharpening of hypotheses for future research, for those methods which do not "impose" a latent model, such as that of LPA. The organization and display features of the proposed strategy may be of greatest significance, especially when stimuli or items are complex and when sorters are heterogenous in terms of how they formed their categories.

Finally, it should be clear that there are several other possible indices of association for comparing matrices of the form of A_s and A_t. Evans (1970) provides one, and others are implicit in his work. Also, S. C. Johnson of Bell Telephone Laboratories has suggested several, including one equivalent to our q_{st}, in an unpublished manuscript ("M[illegible]ic Clustering", circa 1970). The latter paper, in fact, which was discovered after completing most of the present work, includes a number of suggestions which anticipated our own (although Johnson does not use his version of our q_{st} extensively). Nevertheless, we suspect that there is more to gain from critical applications of the available methods than from continued comparison and refinement of methods which, to date, have only rarely been used.

TABLE 1

Partitions of Alphabetic -- Morse Code Characters -- Organized with Target I

CAT	TRACE	H(A)	SRTR																										
5.	.10154	.18462	1.	4.	4.	4.	4.	2.	2.	2.	3.	3.	1.	1.	1.	1.	1.	1.	5.	3.	5.	3.	3.	3.	3.	5.	5.	5.	5.
5.	.10154	.19692	27.	1.	1.	1.	1.	3.	3.	3.	2.	2.	2.	2.	2.	2.	2.	2.	5.	5.	5.	5.	5.	5.	5.	4.	4.	4.	4.
5.	.17846	.23692	40.	1.	5.	2.	3.	4.	4.	4.	1.	1.	1.	1.	1.	1.	1.	1.	2.	2.	2.	2.	2.	2.	2.	3.	3.	3.	3.
5.	.21231	.21538	32.	1.	1.	1.	1.	2.	2.	2.	4.	4.	4.	3.	4.	4.	5.	5.	4.	3.	4.	5.	3.	3.	4.	4.	5.	5.	4.
5.	.24000	.16923	50.	1.	2.	2.	4.	3.	3.	4.	1.	3.	2.	3.	5.	2.	4.	5.	1.	4.	4.	1.	3.	5.	5.	1.	2.	3.	5.
5.	.24615	.16923	44.	1.	3.	1.	5.	3.	1.	5.	1.	3.	4.	3.	2.	3.	5.	2.	4.	5.	1.	4.	5.	2.	2.	4.	4.	3.	2.
5.	.26154	.21538	38.	2.	3.	2.	1.	3.	1.	3.	2.	4.	3.	4.	5.	4.	5.	4.	3.	4.	4.	3.	4.	5.	5.	3.	3.	3.	[illegible]
6.	.11385	.17231	41.	1.	1.	1.	1.	2.	2.	2.	3.	3.	4.	4.	4.	5.	5.	6.	4.	4.	4.	5.	5.	6.	6.	5.	5.	5.	5.
6.	.15077	.14769	37.	2.	2.	2.	2.	1.	1.	1.	3.	5.	6.	6.	3.	3.	6.	6.	5.	3.	4.	5.	3.	5.	5.	5.	4.	3.	4.
6.	.16000	.13846	3.	2.	2.	2.	3.	3.	3.	3.	1.	5.	4.	4.	6.	6.	4.	4.	5.	1.	2.	5.	1.	5.	5.	2.	1.	1.	6.
6.	.16615	.14962	24.	1.	1.	1.	1.	2.	2.	2.	3.	4.	[illegible]	6.	5.	5.	6.	6.	4.	5.	3.	6.	5.	6.	4.	4.	5.	5.	3.
6.	.16923	.14769	14.	1.	1.	1.	1.	2.	2.	2.	3.	4.	6.	4.	3.	3.	6.	4.	5.	3.	5.	4.	3.	4.	6.	4.	5.	3.	5.
6.	.18462	.15692	27.	4.	4.	4.	4.	5.	5.	5.	3.	2.	1.	2.	3.	3.	1.	2.	2.	3.	6.	2.	3.	2.	1.	2.	6.	3.	6.
6.	.18769	.15385	28.	6.	2.	1.	4.	1.	4.	6.	1.	1.	4.	5.	5.	2.	3.	3.	4.	4.	5.	2.	3.	3.	3.	2.	2.	2.	3.
6.	.21538	.19385	43.	1.	1.	1.	1.	2.	2.	2.	4.	3.	5.	3.	4.	4.	5.	3.	3.	4.	3.	3.	4.	3.	5.	3.	6.	4.	3.
6.	.21846	.19154	36.	1.	2.	2.	4.	3.	3.	4.	1.	3.	2.	3.	5.	2.	4.	6.	1.	4.	5.	1.	4.	6.	6.	1.	2.	3.	5.
6.	.27077	.18769	25.	1.	5.	3.	5.	4.	6.	2.	1.	2.	4.	2.	1.	1.	4.	2.	2.	1.	3.	2.	1.	2.	4.	2.	3.	1.	5.
7.	.08308	.12308	9.	1.	1.	1.	1.	2.	2.	2.	3.	3.	4.	4.	3.	3.	7.	7.	3.	4.	4.	5.	5.	5.	5.	3.	6.	6.	6.
7.	.08923	.14154	17.	7.	7.	7.	7.	5.	5.	5.	3.	6.	4.	4.	4.	1.	1.	1.	6.	2.	6.	4.	4.	4.	4.	6.	2.	2.	6.
7.	.14154	.11385	23.	1.	1.	1.	1.	4.	4.	4.	2.	5.	6.	7.	2.	2.	6.	5.	5.	7.	3.	5.	7.	7.	6.	5.	3.	2.	3.
7.	.14462	.12308	26.	3.	3.	3.	3.	4.	4.	4.	2.	5.	7.	5.	2.	2.	7.	5.	6.	1.	6.	5.	2.	5.	7.	5.	6.	1.	6.
7.	.14769	.12000	11.	1.	1.	1.	1.	3.	5.	3.	6.	2.	3.	7.	4.	5.	4.	7.	2.	4.	6.	7.	4.	7.	7.	2.	6.	4.	6.
7.	.14769	.12000	20.	1.	1.	1.	1.	2.	2.	2.	3.	5.	7.	6.	3.	3.	7.	6.	5.	3.	4.	6.	3.	7.	7.	5.	4.	3.	4.
7.	.14769	.13846	5.	1.	1.	1.	1.	1.	1.	1.	2.	3.	5.	6.	2.	2.	7.	7.	3.	6.	4.	6.	6.	6.	5.	3.	7.	7.	4.
7.	.15385	.12615	45.	1.	1.	1.	1.	6.	6.	6.	5.	3.	3.	7.	5.	5.	2.	2.	3.	4.	5.	2.	4.	7.	3.	3.	2.	2.	5.
7.	.18462	.15077	6.	1.	1.	1.	1.	2.	2.	2.	3.	4.	6.	4.	3.	3.	6.	4.	4.	3.	5.	4.	3.	4.	6.	4.	5.	3.	7.
7.	.20923	.12615	4.	6.	3.	7.	6.	5.	1.	4.	6.	4.	5.	1.	6.	2.	5.	1.	4.	2.	7.	1.	2.	4.	2.	4.	3.	6.	2.
8.	.00000	.09538	30.	6.	6.	6.	6.	8.	8.	8.	5.	5.	4.	4.	4.	2.	2.	2.	1.	1.	1.	7.	7.	7.	7.	3.	3.	3.	3.
8.	.00000	.09538	42.	1.	1.	1.	1.	2.	2.	2.	3.	3.	5.	5.	5.	8.	8.	8.	4.	4.	4.	6.	6.	6.	6.	7.	7.	7.	7.
8.	.00000	.0953	46.	3.	3.	3.	3.	5.	5.	5.	1.	1.	2.	2.	2.	7.	7.	7.	6.	6.	6.	4.	4.	4.	4.	8.	8.	6.	6.
8.	.00000	.09538	48.	8.	8.	8.	8.	7.	7.	7.	6.	6.	4.	4.	4.	3.	3.	3.	5.	5.	5.	1.	1.	1.	1.	2.	2.	2.	2.
8.	.00000	.09538	18.	6.	6.	6.	6.	5.	5.	5.	3.	3.	2.	2.	2.	4.	4.	4.	1.	1.	1.	8.	8.	8.	8.	7.	7.	7.	7.
8.	.01538	.09846	47.	1.	1.	1.	1.	6.	6.	6.	2.	2.	3.	3.	3.	6.	5.	5.	8.	8.	8.	7.	7.	7.	7.	4.	4.	4.	4.
8.	.03077	.09538	16.	4.	1.	1.	1.	2.	2.	2.	3.	3.	5.	5.	5.	8.	8.	8.	1.	4.	4.	7.	7.	7.	7.	6.	6.	6.	6.
8.	.12000	.09231	49.	1.	1.	1.	1.	2.	2.	2.	3.	6.	7.	8.	3.	3.	7.	8.	6.	4.	5.	8.	4.	8.	7.	6.	5.	4.	5.
8.	.12923	.09538	35.	5.	5.	5.	5.	1.	1.	1.	8.	4.	7.	4.	6.	6.	7.	4.	2.	3.	8.	4.	3.	2.	7.	2.	8.	3.	6.
8.	.13231	.09846	34.	7.	7.	7.	7.	5.	5.	5.	3.	6.	8.	4.	3.	3.	8.	4.	6.	2.	3.	4.	1.	1.	8.	6.	1.	2.	3.
8.	.13231	.09846	13.	1.	1.	1.	1.	8.	8.	8.	5.	7.	6.	3.	5.	5.	7.	3.	7.	2.	4.	3.	2.	7.	6.	7.	4.	2.	4.
8.	.14462	.09231	31.	2.	2.	7.	2.	3.	3.	3.	7.	6.	1.	5.	7.	8.	1.	5.	6.	7.	4.	5.	8.	6.	1.	6.	4.	8.	4.
9.	.01231	.08308	39.	1.	1.	1.	1.	3.	3.	3.	8.	8.	4.	4.	4.	6.	6.	6.	5.	5.	5.	2.	2.	2.	2.	7.	9.	9.	7.
9.	.02462	.08308	7.	1.	1.	1.	1.	2.	2.	2.	3.	3.	4.	4.	4.	8.	8.	8.	5.	3.	5.	9.	9.	9.	9.	7.	6.	6.	7.
9.	.04923	.09538	8.	2.	2.	2.	2.	5.	3.	9.	9.	9.	5.	5.	5.	3.	3.	3.	6.	7.	6.	8.	8.	8.	8.	4.	4.	4.	1.
9.	.09231	.08308	15.	2.	2.	2.	2.	1.	1.	1.	4.	3.	5.	8.	4.	4.	6.	6.	9.	7.	3.	7.	7.	8.	5.	3.	9.	9.	9.
9.	.09846	.08923	29.	3.	3.	3.	3.	2.	2.	2.	7.	1.	5.	6.	8.	8.	4.	4.	1.	6.	7.	5.	6.	6.	6.	1.	9.	9.	7.
9.	.10154	.08615	10.	6.	8.	8.	8.	7.	7.	7.	9.	6.	4.	2.	9.	9.	5.	1.	6.	5.	3.	2.	5.	2.	4.	6.	3.	3.	3.
9.	.10462	.08308	33.	7.	7.	7.	7.	1.	1.	1.	8.	9.	3.	4.	8.	8.	2.	2.	9.	6.	4.	6.	5.	5.	3.	9.	2.	2.	4.
9.	.11385	.10462	12.	2.	2.	2.	2.	4.	4.	4.	3.	9.	7.	7.	8.	8.	1.	1.	7.	9.	3.	1.	6.	1.	1.	1.	5.	9.	3.
9.	.11692	.08923	21.	3.	3.	3.	3.	4.	4.	4.	7.	1.	8.	6.	2.	2.	9.	9.	1.	5.	7.	9.	5.	6.	8.	1.	9.	9.	7.
9.	.17231	.12615	19.	4.	8.	1.	6.	7.	9.	3.	1.	7.	9.	9.	6.	8.	2.	2.	9.	6.	5.	2.	8.	2.	2.	2.	8.	8.	8.
9.	.18154	.09846	2.	2.	4.	7.	1.	3.	5.	8.	1.	3.	5.	6.	7.	9.	5.	8.	5.	1.	4.	5.	2.	6.	6.	5.	9.	7.	7.
				E	H	I	S	M	O	T	A	N	G	K	W	J	Q	Y	D	R	U	C	P	X	Z	B	F	L	V

TABLE 2

Contingency Table Comparing A PRIORI & Derived Targets for Morse Code Data

A PRIORI Target

LPA Target	1	2	3	4	5	6	7	8
1.	e . i .. s ... h							
2.		t - m -- o ---						
3.			a .-	w .--	j .---			
4.				g --.				
5.				k -.-	q --.- y -.--			
6.						u ..-		
7.						r .-.	c -.-. p .--. x -..- z --..	
8.			n -.			d -..		b -...
9.								f ..-. l .-.. v ...-

TABLE 3

Partitions of Alphabetic -- Morse Code Characters -- Organized with Target II

CAT	TRACE	H(A)	SRTR																										
5.	.11385	.18462	1.	1.	4.	4.	4.	4.	2.	2.	2.	3.	1.	1.	5.	5.	3.	3.	3.	3.	3.	3.	1.	1.	1.	5.	5.	5.	5.
5.	.12000	.19692	22.	2.	1.	1.	1.	1.	3.	3.	3.	2.	2.	2.	4.	5.	2.	5.	5.	5.	5.	5.	2.	2.	2.	4.	4.	4.	5.
5.	.18769	.21538	32.	4.	1.	1.	1.	1.	2.	2.	2.	4.	4.	4.	4.	4.	4.	5.	3.	3.	3.	4.	3.	5.	5.	5.	5.	4.	4.
5.	.19692	.23692	40.	1.	1.	5.	2.	3.	4.	4.	4.	1.	1.	1.	3.	2.	1.	2.	2.	2.	2.	2.	1.	1.	1.	3.	3.	3.	2.
5.	.24000	.16923	44.	4.	1.	3.	1.	5.	3.	1.	5.	1.	3.	2.	4.	4.	3.	4.	5.	5.	2.	2.	3.	5.	2.	4.	3.	2.	1.
5.	.24000	.16923	50.	2.	1.	2.	2.	4.	3.	3.	4.	1.	2.	5.	1.	1.	3.	1.	3.	4.	5.	5.	3.	4.	5.	2.	3.	5.	4.
5.	.26769	.21538	38.	3.	2.	3.	2.	1.	3.	1.	3.	2.	4.	5.	3.	3.	4.	3.	4.	4.	5.	5.	4.	5.	4.	3.	3.	2.	4.
6.	.10154	.14769	37.	6.	2.	2.	2.	2.	1.	1.	1.	3.	3.	3.	5.	5.	5.	5.	3.	3.	5.	5.	6.	6.	6.	4.	3.	4.	4.
6.	.12308	.14462	24.	4.	1.	1.	1.	1.	2.	2.	2.	3.	5.	5.	4.	4.	4.	6.	5.	5.	6.	4.	6.	6.	6.	5.	5.	3.	3.
6.	.13538	.13846	3.	4.	2.	2.	2.	3.	3.	3.	3.	1.	6.	6.	2.	5.	5.	5.	1.	1.	5.	5.	4.	4.	4.	1.	1.	6.	2.
6.	.13538	.15692	27.	1.	4.	4.	4.	4.	5.	5.	5.	3.	3.	3.	2.	2.	2.	2.	3.	3.	2.	1.	2.	1.	2.	6.	3.	6.	6.
6.	.13846	.14769	14.	6.	1.	1.	1.	1.	2.	2.	2.	3.	3.	3.	4.	5.	4.	4.	3.	3.	4.	6.	4.	6.	4.	5.	3.	5.	5.
6.	.17846	.19385	43.	5.	1.	1.	1.	1.	2.	2.	2.	4.	4.	4.	3.	3.	3.	3.	4.	4.	3.	5.	3.	5.	3.	6.	4.	3.	3.
6.	.18154	.17231	41.	4.	1.	1.	1.	1.	2.	2.	2.	3.	5.	4.	5.	4.	3.	5.	5.	4.	6.	6.	4.	5.	6.	5.	5.	5.	4.
6.	.20615	.14154	36.	2.	1.	2.	2.	4.	3.	3.	4.	1.	2.	5.	1.	1.	3.	1.	4.	4.	6.	6.	3.	4.	6.	2.	3.	5.	5.
6.	.21846	.15385	28.	4.	6.	2.	1.	4.	1.	4.	6.	1.	2.	5.	2.	4.	1.	2.	3.	4.	3.	3.	5.	3.	3.	2.	2.	3.	5.
6.	.22154	.18769	25.	4.	1.	5.	3.	5.	4.	6.	2.	1.	1.	1.	2.	2.	2.	2.	1.	1.	2.	4.	2.	4.	2.	3.	1.	5.	3.
7.	.06462	.12308	9.	4.	1.	1.	1.	1.	2.	2.	2.	3.	3.	3.	3.	3.	3.	5.	5.	4.	5.	5.	4.	7.	7.	6.	6.	6.	4.
7.	.09231	.11385	23.	6.	1.	1.	1.	1.	4.	4.	4.	2.	2.	2.	5.	5.	5.	5.	7.	7.	7.	6.	7.	6.	5.	3.	2.	3.	3.
7.	.09231	.13846	5.	5.	1.	1.	1.	1.	1.	1.	1.	2.	2.	2.	3.	3.	3.	6.	6.	6.	6.	5.	6.	7.	7.	7.	7.	4.	4.
7.	.09846	.12000	20.	7.	1.	1.	1.	1.	2.	2.	2.	3.	3.	3.	5.	5.	5.	6.	3.	3.	7.	7.	6.	7.	6.	4.	3.	4.	4.
7.	.11077	.12615	45.	3.	1.	1.	1.	1.	6.	6.	6.	5.	5.	5.	3.	3.	3.	2.	4.	4.	7.	3.	7.	2.	2.	2.	2.	5.	5.
7.	.11385	.14154	17.	4.	7.	7.	7.	7.	5.	5.	5.	3.	1.	4.	6.	6.	6.	4.	4.	2.	4.	4.	4.	1.	1.	2.	2.	6.	6.
7.	.11692	.12000	11.	3.	1.	1.	1.	1.	3.	5.	3.	6.	5.	4.	2.	2.	2.	7.	4.	4.	7.	7.	7.	4.	7.	6.	4.	6.	6.
7.	.12000	.12308	26.	7.	3.	3.	3.	3.	4.	4.	4.	2.	2.	2.	5.	6.	5.	5.	2.	1.	5.	7.	5.	7.	5.	6.	1.	6.	6.
7.	.13538	.15077	6.	6.	1.	1.	1.	1.	2.	2.	2.	3.	3.	3.	4.	4.	4.	4.	3.	3.	4.	6.	4.	6.	4.	5.	3.	7.	5.
7.	.16615	.12615	4.	5.	6.	3.	7.	6.	5.	1.	4.	6.	2.	6.	4.	4.	4.	1.	2.	2.	4.	2.	1.	5.	1.	3.	6.	2.	7.
8.	.07077	.09231	49.	7.	1.	1.	1.	1.	2.	2.	2.	3.	3.	3.	6.	6.	6.	8.	4.	4.	8.	7.	8.	7.	8.	5.	4.	5.	5.
8.	.07385	.09538	42.	5.	1.	1.	1.	1.	2.	2.	2.	3.	8.	5.	7.	4.	3.	6.	6.	4.	6.	6.	5.	8.	8.	7.	7.	7.	4.
8.	.07385	.09538	46.	2.	3.	3.	3.	3.	5.	5.	5.	1.	7.	2.	8.	6.	1.	4.	4.	6.	4.	4.	2.	7.	7.	8.	8.	8.	6.
8.	.07385	.09538	48.	4.	8.	8.	8.	8.	7.	7.	7.	6.	3.	4.	2.	5.	6.	1.	1.	5.	1.	1.	4.	3.	3.	2.	2.	2.	5.
8.	.07385	.09538	18.	2.	6.	6.	6.	6.	5.	5.	5.	3.	4.	2.	7.	1.	3.	8.	8.	1.	8.	8.	2.	4.	4.	7.	7.	7.	1.
8.	.07385	.09538	30.	4.	6.	6.	6.	6.	8.	8.	8.	5.	2.	4.	3.	1.	5.	7.	7.	1.	7.	7.	4.	2.	2.	3.	3.	3.	1.
8.	.07692	.09846	47.	3.	1.	1.	1.	1.	6.	6.	6.	2.	6.	3.	4.	8.	2.	7.	7.	8.	7.	7.	3.	5.	5.	4.	4.	4.	8.
8.	.08308	.09846	13.	6.	1.	1.	1.	1.	8.	8.	8.	5.	5.	5.	7.	7.	7.	3.	2.	2.	7.	6.	3.	7.	3.	4.	2.	4.	4.
8.	.08923	.09846	34.	8.	7.	7.	7.	7.	5.	5.	5.	3.	3.	3.	6.	6.	6.	4.	1.	2.	1.	8.	4.	8.	4.	1.	2.	3.	3.
8.	.09231	.09538	16.	5.	4.	1.	1.	1.	2.	2.	2.	3.	3.	5.	6.	1.	3.	7.	7.	4.	7.	7.	5.	8.	8.	6.	6.	6.	4.
8.	.10462	.09538	35.	7.	5.	5.	5.	5.	1.	1.	1.	8.	6.	6.	2.	2.	4.	4.	3.	3.	2.	7.	4.	7.	4.	8.	3.	8.	2.
8.	.11385	.09231	31.	1.	2.	2.	7.	2.	3.	3.	3.	7.	8.	7.	6.	6.	6.	5.	8.	7.	6.	1.	5.	1.	5.	4.	8.	4.	4.
9.	.05538	.08923	29.	5.	3.	3.	3.	3.	2.	2.	2.	7.	8.	8.	1.	1.	1.	5.	6.	6.	6.	6.	6.	4.	4.	9.	9.	7.	7.
9.	.05538	.08308	15.	5.	2.	2.	2.	2.	1.	1.	1.	4.	4.	4.	3.	9.	3.	7.	7.	7.	8.	5.	8.	6.	6.	9.	9.	9.	3.
9.	.05846	.08615	10.	4.	8.	8.	8.	8.	7.	7.	7.	9.	9.	9.	6.	6.	6.	2.	5.	5.	2.	4.	7.	5.	1.	3.	3.	3.	3.
9.	.06154	.08308	33.	3.	7.	7.	7.	7.	1.	1.	1.	8.	8.	8.	9.	9.	9.	6.	5.	6.	5.	3.	6.	2.	2.	2.	2.	4.	4.
9.	.07385	.08308	39.	4.	1.	1.	1.	1.	3.	3.	3.	8.	6.	4.	7.	5.	8.	2.	2.	5.	2.	2.	4.	6.	6.	9.	9.	7.	5.
9.	.07385	.08308	7.	4.	1.	1.	1.	1.	2.	2.	2.	3.	8.	4.	7.	5.	3.	9.	9.	3.	9.	9.	4.	8.	8.	6.	6.	7.	5.
9.	.08615	.08923	21.	8.	3.	3.	3.	3.	4.	4.	4.	7.	2.	2.	1.	1.	1.	9.	5.	5.	6.	8.	4.	9.	9.	9.	9.	7.	7.
9.	.10462	.09538	8.	5.	2.	2.	2.	2.	5.	3.	9.	9.	3.	5.	4.	6.	9.	8.	8.	7.	8.	8.	5.	3.	3.	4.	4.	1.	6.
9.	.11385	.10462	12.	7.	2.	2.	2.	2.	4.	4.	4.	3.	8.	8.	1.	7.	9.	1.	6.	9.	1.	1.	7.	1.	1.	5.	9.	3.	3.
9.	.17538	.09846	2.	5.	2.	4.	7.	1.	3.	5.	8.	1.	9.	7.	5.	5.	3.	5.	2.	1.	6.	6.	6.	5.	8.	9.	7.	7.	4.
9.	.17846	.12615	19.	9.	4.	8.	1.	6.	7.	9.	3.	1.	8.	6.	2.	9.	7.	2.	8.	6.	2.	2.	9.	2.	2.	8.	8.	8.	5.
				G	E	H	I	S	M	O	T	A	J	W	B	D	N	C	P	R	X	Z	K	Q	Y	F	L	V	U

Table 4 a

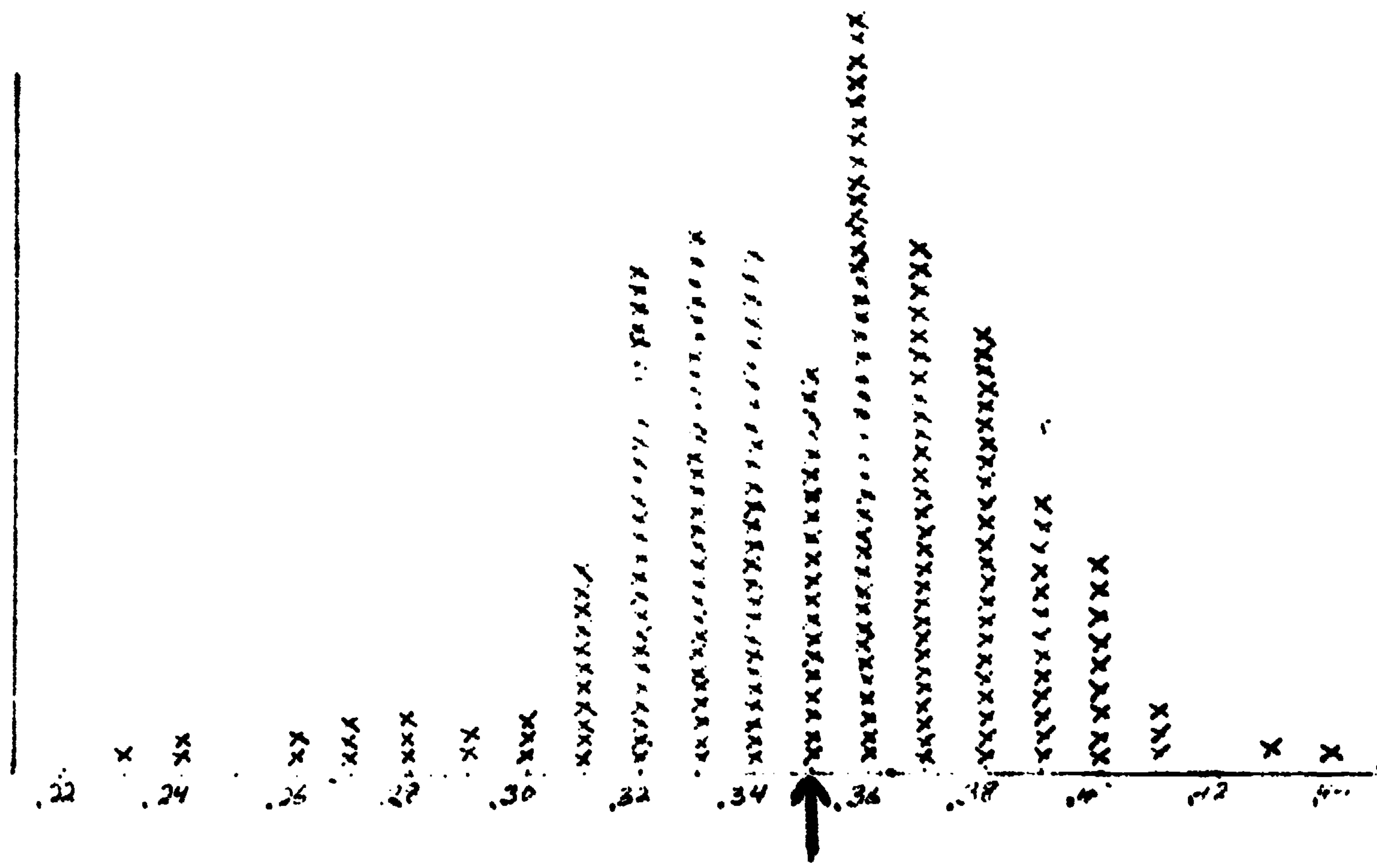
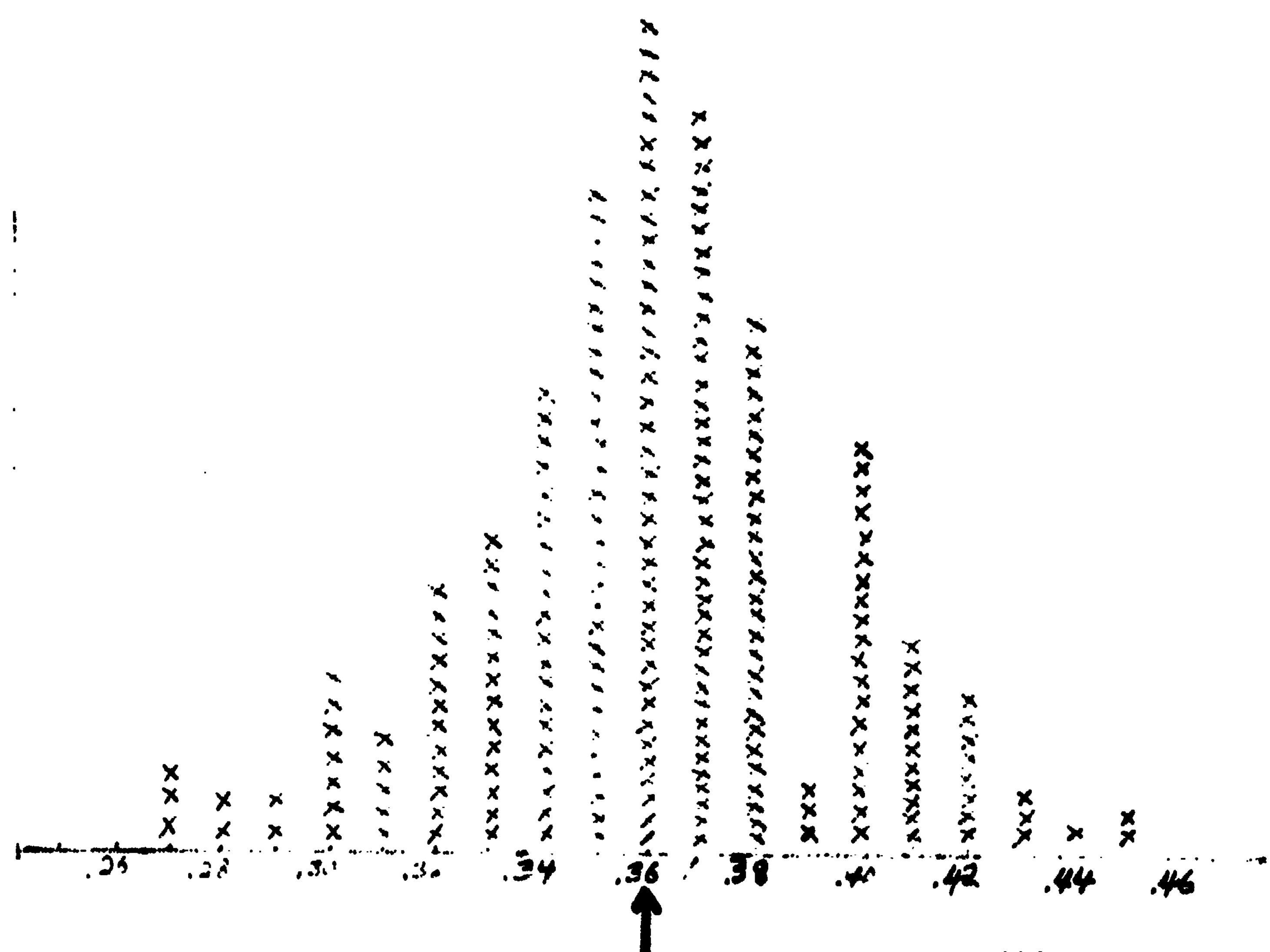


Table 4 b



Monte carlo distributions for 250 random partitions of the statistic q_{st}; N = 20 & $n_s = n_t = 4$ for each

REFERENCES

BENNETT, J. F., and HAYS, W. L. (1960). Multidimensional unfolding: determining the dimensionality of ranked preference data. Psychometrika, 25, 27-43.

COOMBS, C. H. (1950). Psychological scaling without a unit of measurement. Psychological Rev., 57, 148-158.

EVANS, G. T. (1970). The analysis of categorizing behavior. Psychometrika, 35, 367-392.

GUTTMAN, L. (1968). A general nonmetric technique for finding the smallest coordinate space for a configuration of points. Psychometrika, 33, to be published.

HARRIS, C. W., and KAISER, H. F. (1964). Oblique factor analytic solutions by orthogonal transformations. Psychometrika, 29, 347-362.

HARTIGAN, J. A. (1967). Representation of similarity matrices by trees. J. Amer. Statist. Assoc., 62, 1140-1158.

HORAN, C. B. (1969). Multidimensional scaling: Combining observations when individuals have different perceptual structures. Psychometrika, 34, 139-165.

JOHNSON, S. C. (1967). Hierarchical clustering schemes. Psychometrika, 32, 241-254.

KRUSKAL, J. B. (1964a). Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. Psychometrika, 29, 1-27.

KRUSKAL, J. B. (1964b). Nonmetric multidimensional scaling: a numerical method. Psychometrika, 29, 115-129.

LANCE, G. N., and WILLIAMS, W. T. (1966/67). A general theory of classificatory sorting strategies. 1. Hierarchical systems. The Computer Journal, 9, 373-380.

MILLER, G. A. (1969). A psychological method to investigate verbal concepts. Jour. of Mathematical Psychology, 6, 169-191.

SHEPARD, R. N. (1962). The analysis of proximities: multidimensional scaling with an unknown distance function, Pts. I and II. Psychometrika, 27, 125-140, 219-246.

SHEPARD, R. N., and CARROLL, J. D. (1966). Parametric representation of nonlinear data structures. In "Multivariate Analysis." (P. R. Krishnaiah, ed.), pp. 561-592. Academic Press, New York.

SOKAL, R. R., and SNEATH, P. H. A. (1963). "Principles of Numerical Taxonomy." Freeman, San Francisco, California.

TRYON, R. C. (1958). General dimensions of individual differences: Cluster analysis vs. multiple factor analysis. Educ. and Psych. Measurement, 18, 477-495.

TUCKER, L. R., and MESSICK, S. (1963). An individual differences model for multidimensional scaling. Psychometrika, 28, 333-367.